A Study on Principal Component Analysis over Wireless Channel

R. Upadhyay, A. Soni, P. Panse and U. R. Bhatt
Department of Electronics & Telecommunicati, Institute of Engineering & Technology,
Devi Ahilya University, Khandwa Road, Indore 452001, India
rupadhyay@ietdavv.edu.in

***Abstract*— Applications in many fields such as the internet of things (IoT), stock market, image compression, food adulteration, wireless physical layer key generation, etc. are becoming progressively complex due to a large number of users and increment in their usage. Data obtained by these applications are in huge amount creating a high computational cost. Further, it is difficult to handle and analyze it. To deal with such problems, dimensionality reduction techniques are used and one of the dimensionality reduction techniques is the Principal Component Analysis (PCA). In this paper, PCA is applied over a wireless Rician channel with AWGN at different SNR. It is concluded that the information content is more in less number of principal components with samples at higher SNR. It is also observed that the different combinations of several groups and elements in the sample space provide a different cumulative percentage of information.**

***Index Terms* — AWGN, Dimensionality Reduction, Principal Component Analysis, Rician Channel.**

I. Introduction

At present, the advancement of technology that generates a huge amount of data has rapidly increased the number of users, leading to problems in handling and processing huge data. These problems can be mitigated by reducing its dimensions. Principal Component Analysis (PCA) can be used as a dimensionality reduction technique on the wireless Rician channel with additive white Gaussian noise (AWGN). PCA sorts data sets in terms of information content for data analysis and processing. Dimensionality reduction also represents variance and provides a low dimensional characterization of data [1].

PCA, along with dimensionality reduction, performs data conditioning. Firstly, principal components are formed such that they are mutually uncorrelated [2]. Principal components are arranged in terms of variability, that is the first principal component has the largest variability among data points, and this decreases until the last principal component [2]. Due to the uniqueness of information, greater variability can also be termed as greater information. Lastly, principal components are mutually orthogonal [2].

In this paper, we studied PCA as a dimensionality reduction technique on RSS samples collected from AWGN channel at different SNR in the presence of fading. We analyzed various RSS samples corresponding to the extracted principal components. It is observed that the total number of principal components required for a fixed amount of information varies with the different combination of row and column of the input data set and can be utilized for dimensionality reduction for many resource constraint applications like IoT.

This paper is arranged as follows: Section II comprises of related work, section III describes the principal component analysis, followed by the proposed preprocessing using PCA in section IV. Section V consists of experimental setup and result, followed by the conclusion in section IV

II. Related Work

PCA has been used in applications like mechanical fault estimation method to estimate the probability of working of mechanical equipment [3], in stock exchange [4], in mobile robots with sensors like catadioptric camera [5]. To determine various unspecified mineral oils, PCA was used by the author in [6]. In [7], the author used PCA in business intelligence, while in [8], the author had applied PCA along with grey relational analysis for the appropriate design of machine parameters in electro-discharge diamond face grinding. Studies also have been carried out over PCA for a variety of wireless applications. The authors in [9] suggested using PCA for enhancing Wi-Fi fingerprinting for the indoor positioning system. PCA was used for noisy measurement removal and compressing the dimensions of received signal strength in online as well as offline phases. Li et al. in [10] applied PCA in secret key establishment for wireless channels. It generated an uncorrelated secret key at both ends of the wireless channel. It also compared signal preprocessing methods, i.e., PCA, discrete cosine transforms (DCT), and wavelet transforms (WT). Among these methods, PCA has proved to be the best as the increased key agreement, decreased information leakage, and reduced computation cost. Also, key generation by using common Eigenvector can give the benefit of a greater key generation rate, reducing the key error rate and good randomness. For wireless physical layer security, the use of frequency hopping along with PCA and moving average, followed by double-threshold quantization can cause a drastic increment in key agreement and key generation efficiency have been presented in [11]. The author in [12] employed PCA as a dimensionality reduction for user information in the universal network, and it is shown that out of the seventeen principal components, nine can give 88.4% information. PCA was used as a dimensionality reduction technique for network intrusion detection in [13]. It is shown that the first ten principal components obtain an accuracy of 98% out of forty-one features extracted. In [14], the author had used PCA for lightening variability in images of Lambertian object. It is shown that the first five principal components give 97% of the information.

From the above-surveyed literature, it is clear that PCA can be applied as a dimensionality reduction technique for a wide range of applications, hence it is analyzed in this work.

Although the existing work determines the number of principal component groups to obtain a specific amount of information, the analysis of the data set, namely the number of elements/groups to extract maximum information from the minimum number of principal components is missing as per our findings and is presented in this work. The selection of the minimum number of principal components with maximum information improves the computational efficiency of PCA as a dimensionality reduction technique. Further, the SNR of the system is improved by applying PCA in the literature [15], but the effect of different SNR on the performance of PCA is untouched. In this work, we applied PCA as a dimensionality reduction technique for an application like physical layer key generation and analyzed the collected RSS samples with the proposed approach to eliminate the redundant information to make PCA efficient.

III. PRINCIPAL COMPONENT ANALYSIS

PCA is a dimensionality reduction technique based on feature selection. PCA works on the principle of transforming high dimensional correlated data sets into low dimensional uncorrelated principal components. PCA can be computed by using two methods: Covariance matrix calculation or singular value decomposition (SVD) [16]. Besides involved in dimensionality transformation, PCA also provides the arrangement of principal components decreasing in terms of variability along with the amount of information from first to last principal component.

PCA transforms the original data sets into principal components such that their values are completely changed. In this case, the coordinates of data are changed when the original coordinate system data is placed and the direction of the largest variability determines the axes of the first principal component. The axes of the second principal component are the second-largest variability direction and it should be orthogonal to the axes of the first principal component. The original data set is linearly combined to obtain the principal component analysis. At the first iteration, it resembles linear regression between any of two axes. But, unlike linear regression, it reduces perpendicular length between the data and principal components besides reducing the distance between the variable and the predicted value [17].

There are some variants of PCA, for example in the Simultaneous Principal Component Analysis (SMPCA), all the eigenvectors corresponding to all eigen values are generated by using the covariance matrix [18]. In Progressive Principal Component Analysis (PGPCA), the first largest eigen value is taken, and then the eigenvector corresponding to that eigen value is found. After that, the orthogonal subspace projection (OSP) operator is applied and the subspace is extracted, causing it to be linearly covered to the first eigenvector that occupies the space. Then, the second largest eigen value is taken, and the procedure is repeated for all the needed eigen values. In the Successive Principal Component Analysis (SCPCA), the first projection vector is randomly generated, in which the first dimension is best fitted in its direction is calculated. It is repeated for the second and the subsequent eigenvectors. In the Prioritized Principal Component Analysis (PRPCA), the part of prioritized PCA is grown. As in the case of SCPCA, the random vectors are not used, but the self-modifying initialization method is used to create the initial projection vectors for PCA. In sequence, the prioritization of principal components from the projection vector is created by the initialization method.

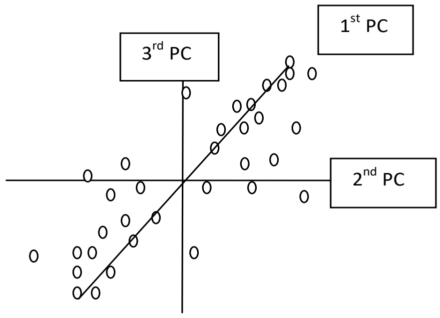


Figure 1: Principal components showing orthogonal property with variance decreasing from first principal component to third.

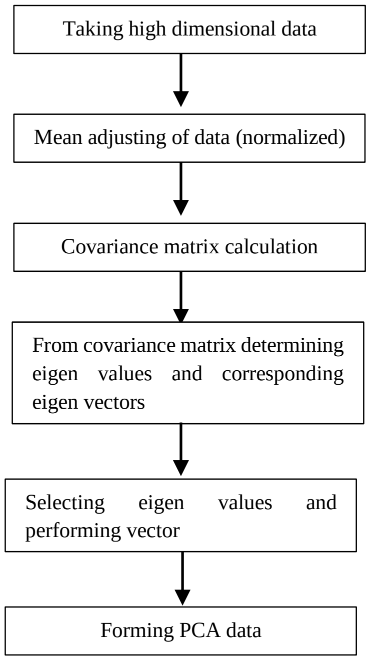


Figure 2: Steps involved in PCA

IV. PROPOSED RSSI PREPROCESSING USING PCA

We have proposed to apply pre-processing PCA over RSSI sample, which was calculated using the covariance matrix calculation method as it is less complex compared to the singular value decomposition (SVD). We have created original data set by collecting RSSI values of the wireless channel and processed it with PCA, so that the dimension of data can be reduced for many resource constraint applications, like IoT and WSN. PCA can also be used for wireless physical layer key generation to improve the performance of the system.

In our proposed model, namely the Simultaneous Principal Component Analysis (SMPCA), the variant of PCA is used because the characteristics of groups along with their percentage of information content need to be observed, so it is necessary to compute all principal components. In this regard, the number of groups and elements that determines the least number of principal components with a maximum percentage of information content is determined. Also, for the analysis of variance in individual groups, it is needed to find all the principal components. Various steps involved in the proposed preprocessing using PCA are shown in Figure 3.

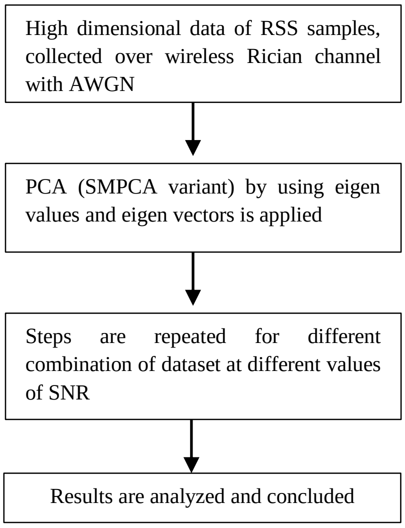


Figure 3: Proposed RSSI preprocessing using PCA

V. EXPERIMENTAL SETUP AND RESULTS

This section describes the experimental setup collecting RSS values processed by PCA. NodeMCU ESP8266 performs bit extraction at the frequency of 2.4 GHz. RSS signals are collected in an indoor static environment (furniture and other surroundings fixed) by using Tx and Rx in wireless Rician channel considering AWGN. One of the equipment was stationary and the other was moving at walking speed. RSS samples were collected alternately at both equipment.

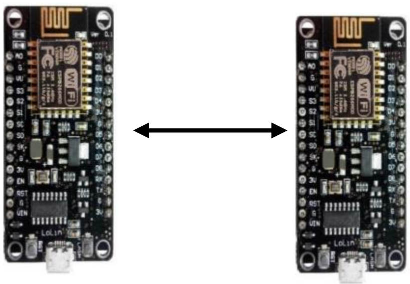


Figure 4: Experimental setup for collecting RSSI values.

To achieve minimum variation between measurements, the time delay of measurements was kept less than the coherence time of the channel. PCA was performed by using Eigen value and Eigen vector with SMPCA. Figure 4 shows an experimental setup and the description of simulation parameters is given in Table 1.

Table 1
Description of experimental setup

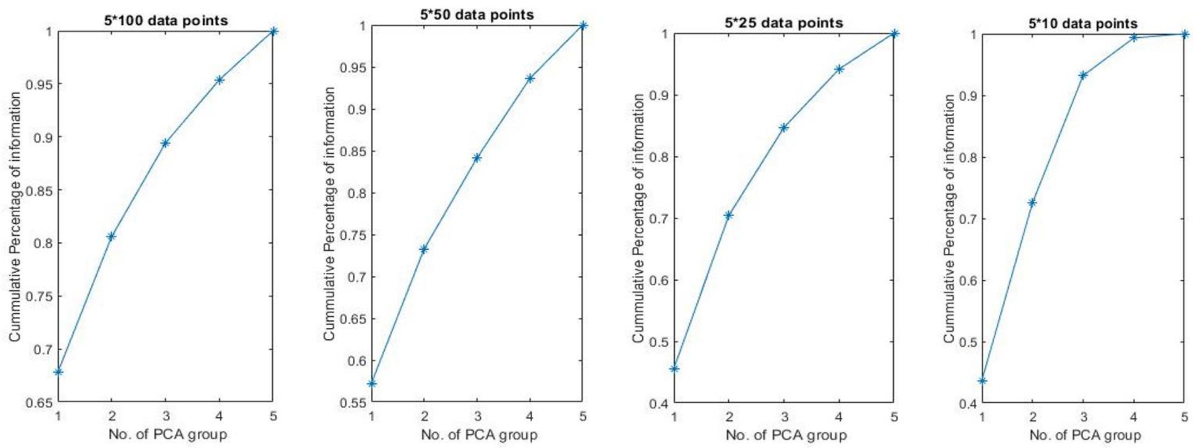
Parameter	Value
Environment	Indoor
Channel	Rician Channel
Type of PCA	using Eigen value and Eigen vector
Subtype of PCA	SMPCA
Distance between Tx and Rx	1 meter

Figure 5 and 6 show the analysis of the cumulative percentage of information with a different number of groups (5,15) and the number of elements (100,50,25,10) in each PCA groups, in the collected RSSI values. Table 2 summarizes the number of principal components comprising the cumulative information of 90 % for different combinations of data sets.

Table 2
Cummulative information corresponding to different dataset

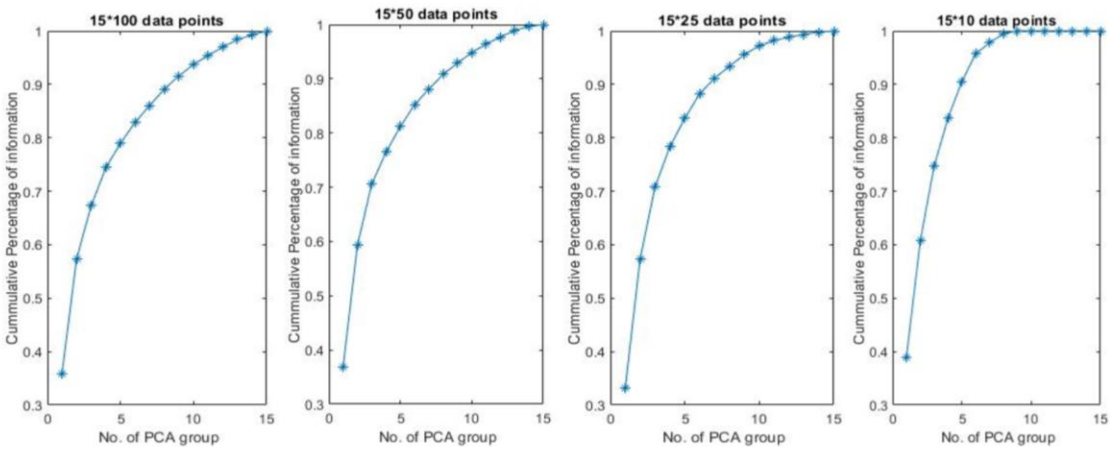
S. No.	Data set dimensions (groups * elements/group)	Total number of Principal Components	Principal Components with 90 % of information	Figure
1.	5*100	5	3	5(a)
2.	15*100	15	8-9	6(a)
3.	5*50	5	3-4	5(b)
4.	15*50	15	8	6(b)
5.	5*25	5	3-4	5(c)
6.	15*25	15	7	6(c)
7.	5*10	5	2-3	5(d)
8.	15*10	15	5	6(d)

From Table 2, it can be observed that the results obtained from Figure 6 (d), that is for dimensions 15*10 is the best. In this case, out of 15 principal components, 90 percent of cumulative information is acquired by the first five components only. It can be seen that when the number of groups is increased, and the elements in these groups are decreased, there will be a higher possibility to eliminate groups with none or the least loss of information. So, dimensionality reduction factor will be increased and PCA can be applied more efficiently.



(a) (b) (c) (d)

Figure 5: Cumulative percentage of information corresponding to data set with 5 groups and: (a) 100 elements/group, (b) 50 elements/group, (c) 25 elements/group, (d) 10 elements/group.



(a) (b) (c) (d)

Figure 6: Cumulative percentage of information corresponding to data set with 15 groups and: (a) 100 elements/group, (b) 50 elements/group, (c) 25 elements/group, (d) 10 elements/group.

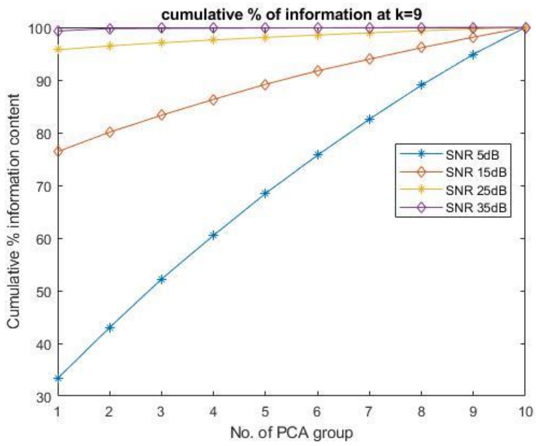


Figure 7: Cumulative information (%) corresponding to 10 principal components over Rician channel at k = 9.

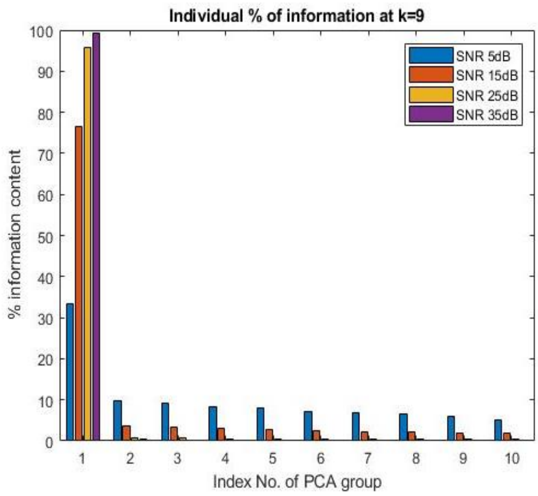


Figure 8: Individual information (%) corresponding to 10 principal components over Rician channel at k=9

Further, Figure 7 and 8 show cumulative and elemental (individual) percentage of information content corresponding to the principal components with different SNR values. It is observed that out of 10 principal components at SNR 5dB, 15dB, 25dB, and 30dB, the required number of principal components to achieve 90% of information are 8-9, 6 and 1 respectively.

PCA carried out with 35dB gives the best result. In other words, the first few principal components contain the majority of the information; thus, it gives the best dimensionality reduction results, and this characteristic decreases with a decrease in the value of SNR.

VI. CONCLUSION

In this paper, cumulative percentage of information between data with a different number of groups and elements was acquired and it is observed that different combination gives different results, in which more number of groups and less number of elements in each group gives the best result. Further, we also examined the cumulative and elemental (individual) percentage of information content corresponding to principal components with different SNR. It is observed that PCA carried out with the highest SNR provides better result in terms of dimensionality reduction. In future, the analysis can be done for more specific applications to achieve the appropriate amount of information content in much least number of Principal components with the fact that each application requires a specific percentage of information to be kept.

REFERENCES

[1] Shruti Sehgal, Harpreet Singh, Mohit Agarwal, V. Bhasker, Shantanu, Data analysis using principal component analysis, IEEE, International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014
[2] Md Shahid Latif, "Principal Component Image Interpretation – A Logical and Statistical Approach'', IJEDR, Volume 2, Issue 4, 2014, ISSN: 2321-9939
[3] Luhui Lin, Jie Ma, Xiulan Ye, Xiaoli Xu Mechanical fault prediction based on principal component analysis, IEEE International Conference on Information and Automation, 2010
[4] Muhammad Waqar, Hassan Dawood,Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, "Prediction of Stock Market by Principal Component Analysis ",Ping Guo,2017 13th International Conference on Computational Intelligence and Security (CIS),2017, 17578286
[5] M. Artac, Mobile robot localisation with incremental PCA, 11th IEEE Mediterranean Electrotechnical Conference, 2002
[6] David A. Rusak, Leah M. Brown, and Scott D. Martin, "Classification of Vegetable Oils by Principal Component Analysis of FTIR Spectra'', Journal of chemical education (*J. Chem. Educ.*), 2003, 80 (5), Pp 541
[7] Ana-Maria Sevcenco ,Kin Fun Li, " Principal Component Analysis in Business Intelligence Applications''2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2013, 13972144
[8] G. K. SinghȦ , N. K. ChauhanḂ , Rajeev KumarĊ and V. YadavaḊ, Grey Relational Analysis Coupled with Principal Component Analysis for Optimization Design of the Machining Parameters in Electro-

Discharge Diamond Face Grinding, International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 – 5161, 2014

[9] Y. Basiouny, M. Arafa,Amany M. Sarhan, “Enhancing Wi-Fi fingerprinting for indoor positioning system using single multiplicative neuron and PCA algorithm,’’ 2017 12th International Conference on Computer Engineering and Systems (ICCES), 2017, 17542818

[10] Guyue Li, Aiqun Hu, Junqing Zhang, Linning Peng, Chen Sun, and Daming Cao, “High-Agreement Uncorrelated Secret Key Generation Based on Principal Component Analysis Preprocessing’’, IEEE Transactions on Communications, Volume: 66, Issue: 7, July 2018, 17916674: Pp 3022 - 3034

[11] 11) Liangjun Hu, Fangyu Zhang, Aiqun Hu, Yu Jiang, Guyue Li, “A key generation scheme for wireless physical layer based on frequency hopping”, 8th International Congress of Information and Communication Technology (ICICT-2018), Volume 131, 2018, Pp 1104-1112

[12] Yu Dai, Jianfeng Guan, Wei Quan, Changqiao Xu, Hongke Zhang, PCA-based dimensionality reduction method for user information in universal network , IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, 2012

[13] K. Keerthi Vasan, B. Surendiran, Dimensionality reduction using Principal Component Analysis for network intrusion detection, Elsevier (Volume 8), September 2016

[14] Ravi Ramamoorthi, Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object, IEEE transactions on pattern analysis and machine intelligence, 2002

[15] Moussa Hamadache, Dongik Lee, improving signal-to-noise ratio (SNR) for inchoate fault detection based on principal component analysis (PCA), 14th International Conference on Control, Automation and Systems, 2014

[16] Liton Chandra Paul, Abdulla Al Suman , Nahid Sultan, Methodological Analysis of Principal Component Analysis (PCA) Method, IJCEM International Journal of Computational Engineering & Management, Vol. 16 Issue 2, March 2013

[17] Hafiz zia ur Rehman , Sungon Lee, Automatic Image Alignment Using Principal Component Analysis, IEEE Access (Volume: 6), 20 November 2018

[18] Liu Wei-min, Chang Chein-I, Variants of Principal Components Analysis, IEEE International Geoscience and Remote Sensing Symposium, 2007.